Referee 2 (review response received and published on 1 Mar. 2018)

The authors would like to thank Referee 2 for his or her comments on the manuscript. The comments and changes are addressed below, where the referee’s original comments are copied in italics. All page and line numbers in the referee’s comments refer to the originally submitted version of the paper.

The authors tested two methods on reducing spatial sampling errors in the comparison of profiles of collocated radiosonde and GPSRO. They concluded that the “double differencing” method (by using a NWP model data) will most effectively reduce the sampling errors. And if that method is not available, restricting the comparisons to within an ellipse parallel to the wind flow will also reduce the sampling errors caused by atmospheric variability. I like this paper. But the authors discussed only the impact on root-mean-square (rms) errors, not mentioning at all the impact of the mean differences between GPSRO and radiosonde. This authors should add plots or tables to indicate how different methods impact the mean difference…that is very important to studies of utilizing RO data to analyze radiosonde accuracy.

Thank you for the suggestion. We have computed the mean and standard deviation differences of the three variables for all stations, and these figures are now included in Appendix A. Our results indicate that the mean is not sensitive to co-location or sampling differences, as all six co-location methods result in similar mean profiles for all variables across all stations. This is most likely due to cancelation of differences of opposite signs during the computation of the mean. However, we see the impact of the different co-location criteria on the standard deviations, which are very similar to the RMS differences. The mean differences also give some indication of the biases in the RO minus radiosonde differences, as discussed in the new Appendix A.
Reducing representativeness and sampling errors in radio occultation–radiosonde comparisons

Shay Gilpin 1, Therese Rieckh 1,2, and Richard Anthes 1
1COSMIC Program Office, University Corporation for Atmospheric Research, Boulder, CO
2Wegener Center for Climate and Global Change, University of Graz, Austria

Correspondence to: Shay Gilpin (sgilpin@ucar.edu)

Abstract. Radio occultation (RO) and radiosonde (RS) comparisons provide a means of analyzing errors associated with both observational systems. Since RO and RS observations are not taken at the exact same time or location, temporal and spatial sampling errors resulting from atmospheric variability can be significant and inhibit error analysis of the observational systems. In addition, the vertical resolutions of RO and RS profiles vary and vertical representativeness errors may also affect the comparison. Typically in RO–RS comparisons, RS observations are co-located with RO profiles within a fixed time window and distance, i.e. within 3–6 h and circles of radii ranging between 100–500 km. In this study, we first show that vertical filtering of RO and RS profiles to a common vertical resolution reduces representativeness errors. We then test two methods of reducing horizontal sampling errors during RO–RS comparisons: restricting co-location pairs to within ellipses oriented along the direction of wind flow rather than circles, and applying a spatial-temporal sampling correction based on model data. Using data from 2011–2014, we compare RO and RS differences at four GCOS Reference Upper-Air Network (GRUAN) RS stations in different climatic locations, in which co-location pairs were constrained to a large circle (~666 km radius), small circle (~300 km radius), and ellipse parallel to the wind direction (~666 km semi-major axis, ~133 km semi-minor axis). We also apply a spatial-temporal sampling correction using European Centre for Medium-Range Weather Forecasts Interim Reanalysis (ERA-Interim) gridded data. Restricting co-locations to within the ellipse reduces root mean square (RMS) refractivity, temperature, and water vapor pressure differences relative to RMS differences within circles of similar area. Applying the spatial sampling correction shows the most significant reduction in RMS differences, such that RMS differences are nearly identical with the spatial sampling correction regardless of the geometric constraints. We conclude that implementing the spatial-temporal sampling correction using a reliable model will most effectively reduce sampling errors during RO–RS comparisons, however, if a reliable model is not available, restricting spatial comparisons to within an ellipse parallel to the wind flow will reduce sampling errors caused by horizontal atmospheric variability.

1 Introduction

Radio occultation (RO), a relatively new method of atmospheric measurements, has established itself as an important atmospheric observational system. By measuring the phase delay of radio-waves sent from Global Positioning Satellites (GPS)
traversing horizontally-quasi-horizontally through Earth’s atmosphere to low-Earth orbiting satellites, RO obtains accurate and precise vertical profiles of bending angles (Melbourne et al., 1994). Refractivity is obtained from by inverting bending angle profiles using the Abel Transform, and from refractivity, stratospheric temperature and tropospheric temperature and water vapor can be retrieved (where tropospheric temperature and water vapor require auxiliary measurements or modeldata). RO provides observations with global coverage, high vertical resolution, little or no instrument bias, and bending angles which can be directly transformed. Refractivity is a function of temperature and water vapor pressure, therefore with auxiliary information (observations or model) of one, the other can be retrieved. Either bending angles or refractivity may be assimilated into numerical weather prediction models (Melbourne et al., 1994; Kursinski et al., 1997; Anthes et al., 2008) (Eyre, 1994).

Since the proof-of-concept GPS/MET mission in 1995 (Ware et al., 1996), RO profiles of refractivity, temperature, and water vapor have been compared to radiosondes (RS) to assess the quality of RO retrievals and the performance of RS instruments. RS are considered a standard for comparison due their long history of in situ measurements starting in 1958 (Ho et al., 2017), and several studies have used RS as a reference for RO retrieval analysis (Wickert et al., 2004; Kuo et al., 2005; He et al., 2009; Xu et al., 2009; Ho et al., 2010; Sun et al., 2010; Zhang et al., 2011; Wang et al., 2013; Vergados et al., 2014). Conversely, due to RO’s properties of high accuracy and precision, high vertical resolution, and global coverage, RO has been used to evaluate the performance of various RS instruments. For example, Kuo et al. (2005) demonstrated that RO observations are of high enough accuracy and resolution to differentiate between RS instruments and assess their performance, particularly instrument biases due to geographic region, radiation errors, day/night biases, etc.

One of the main difficulties associated with RO–RS comparisons comes from temporal and spatial differences between nearby RS and RO soundings. Since both measurements are not taken at the exact same time or location, temporal and spatial errors (sampling errors) can be a significant part of the computed RO–RS differences. To reduce the effects of sampling errors, the majority of previous studies have restricted co-located RO observations to within a fixed time range and distance, typically within 3 – 6 h of the RS launch and within circles of radii ranging from 100 – 500 km centered at the RS launch site. Alternatively, Staten and Reichler (2009) used smaller radii circles (3 – 36 km) and fitted a second order polynomial to the root mean square (RMS) differences in order to filter out atmospheric variability. Weather-scale atmospheric variability within these circles and time ranges is the major cause of these sampling errors (Sun et al., 2010), and in this study we...

Kitchen (1989) compared RS with infrared (IR) soundings and noted that sampling errors generally dominate the total error associated with the comparison. Bruce et al. (1977) compared RS with satellite retrieved temperature, also discussing the impacts of sampling errors. Mears et al. (2015) applied a bilinear best fit plane to microwave radiometer observations to reduce sampling errors during comparisons with ground-based GPS observations. Fasso et al. (2014) and Ignaccolo et al. (2015) both propose a statistical modeling approach to reduce sampling errors associated with RS inter-comparisons and balloon drift analysis. In this study, we focus on RO–RS comparisons, and attempt to reduce these errors the sampling errors that occur during these comparisons.

We apply two methods to reduce sampling errors caused by atmospheric variability in RO–RS comparisons. First, we restrict co-location pairs to within ellipses oriented along the direction of wind flow rather than circles. Orienting ellipses along the direction of wind flow should minimize atmospheric variability due to the interdependence of atmospheric composition and
Wind. Temperature and water vapor gradients in the free atmosphere tend to be perpendicular to wind flow, resulting in refractivity (a function of both temperature and water vapor pressure) gradients to also approximately lie perpendicular to wind flow. Therefore, we hypothesize that the spatial variability of refractivity within ellipses of semi-major axis $a$ oriented along the wind direction will be reduced compared to the variability within circles of radius $a$. The second method consists of applying a spatial-spatial-temporal sampling correction to the RO–RS co-location pairs based on using the “double-difference” method (Chander et al., 2013; Tradowsky et al., 2017). In the double-difference method, each data set is compared to an intermediate reference data set; in our study we use the European Centre for Medium-Range Weather Forecasts Interim Reanalysis (ERA-Interim) model data (see www.ecmwf.int/era). (Dee et al., 2011, and https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim).

By subtracting the model data from both the reference model data at the corresponding RO and RS locations in space and time from the RO and RS observations, the spatial-temporal sampling errors are largely eliminated, leaving mainly the background and representativeness errors are removed, leaving only the RO and RS observational errors. The approach has been called the “double-difference” method (Haimberger et al., 2012; Wong et al., 2015; Tradowsky et al., 2017). In addition to the sampling error, the different vertical resolution of the RO and RS profiles can lend to representativeness errors when compared. To reduce these errors, we filter the profiles to a common vertical resolution before the comparisons.

Although this paper considers RO–RS comparisons, the vertical filtering and methods of reducing sampling errors can be applied to comparisons of other data pairs, such as any two sounding systems or observations and models. However, the amount of filtering and the geometric constraints on the co-location pairs may have to be adjusted for different comparisons. For example, comparisons of RO or RS soundings with infrared (IR) or microwave (MW) soundings, which have much different vertical resolutions, would require a greater filtering of the RO or RS profiles to make them comparable to the lower-resolution profiles.

The first section describes the data sets, filtering methods, and methodology implemented in this study. Next, we discuss aspects of the ellipse co-location method conducted using the Lindenberg RS station. The following section describes the results of co-locations using both the ellipse and spatial-sampling correction methods at four different RS stations. In the final section we summarize the results and discuss further impacts and implications, followed by an appendix which includes mean and standard deviation differences and further discussion of the spatial-temporal sampling correction.

2 Methods

2.1 RO and RS data sets

All RO profiles are provided by the COSMIC Data Analysis and Archive Center (CDAAC), which can be found at http://cdaac-www.cosmic.ucar.edu. RO refractivity, temperature, and water vapor pressure profiles are taken from the wetPrf profiles provided on a uniform 100 m mean sea level height grid. All available processed, high-quality profiles from the COSMIC-1, GRACE 1 and 2, and Metop-A/B missions during the time periods of comparisons are used. Summaries of these missions can be found at the CDAAC website above. The temperature and water vapor pressure profiles are computed using a 1D-VAR

Figure 1. Map of the GRUAN RS network. Labeled locations with red dots are RS stations used in this study, black dots mark the other GRUAN network RS stations. For more information about these and other GRUAN stations, see https://www.gruan.org.

All RS profiles are provided by the Global Climate Observing System (GCOS) Reference Upper-Air Network (GRUAN; see Bodeker et al., 2016, for more information on the GRUAN project) and downloaded from the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center. GRUAN RS were chosen for their reference-quality observations and well documented uncertainties (Seidel et al., 2009; Immler et al., 2010; Seidel et al., 2011), allowing for a better analysis of the reduction of sampling errors associated with each co-location method.

We chose four stations in different climates for this study: Lindenberg, Germany (LIN); Ny-Ålesund, Norway (NYA); Tateno, Japan (TAT); and Nauru, Nauru (NAU) for the following time periods: 2014, 2013, 2012, and 2011–2013, respectively. (Nauru is the only station in which the full period of activation was used; this is due to the low number of RS launches during 2011 through late August 2013.) Figure 1 is a map of the GRUAN RS stations with the locations of the four stations used in this study labeled and marked in red. All four stations use the Vaisala RS92 radiosonde instruments (for details on GRUAN RS processing see Dirksen et al., 2014).

For refractivity comparisons, RS refractivity is computed under the assumption of a neutral atmosphere (Smith and Weintraub, 1953):

\[ N = 77.6 \frac{P}{T} + 3.73 \times 10^5 \frac{e}{T^2} \]  

where \( N \) is refractivity (N units), \( \frac{P}{T} \) is dry air pressure (hPa), \( T \) is temperature (Kelvin) and \( e \) is water vapor pressure (hPa).

We generate two data sets: an unfiltered data set which contains all original RO and RS profiles, and a filtered data set containing the vertically filtered versions of the original RO and RS profiles.
2.2 Filtering

Representativeness errors result from two different aspects of the RO and RS observations: (1) GRUAN RS have a temporal resolution of 1 s, with vertical resolution of 5–10 m on average (Ladstädter et al., 2015), which is much finer than the 100 m vertical resolution of the RO wetPrf profiles, and (2) RS observations are a series of point measurements, whereas RO observations are weighted averages of a cylindrical volume of atmosphere with horizontal scales of 150–300 km (Kursinski et al., 1997; Kuo et al., 2004; Anthes, 2011). Vertical filtering of both RO and RS profiles should decrease the representativeness errors caused by differences in vertical resolutions and observation type.

Vertical filtering of both RO and RS profiles should decrease the representativeness errors caused by differences in vertical resolutions and observation type (Lohmann, 2007), resulting in a more meaningful comparison between profiles. Removal of structures with very small vertical scales also has the effect of reducing representativeness errors associated with different horizontal scales (footprints) of the observations (Kitchen, 1989). Structures with very small vertical scales in the RS profile are likely associated with horizontal scales too small to be resolved by the RO observations. Though the majority of previous RO–RS comparison studies do not filter the RO or RS profiles before comparison, Kuo et al. (2004) filtered the profiles to remove structures with vertical scales less than 1 km before comparison with model analyses in an effort to minimize vertical sampling differences.

To remove small-scale, unrepresentative structures in both the RO and RS profiles, we applied the Savitzky-Golay low-pass filter (Savitzky and Golay, 1964). We first linearly interpolated profile variables (refractivity, temperature, and water vapor pressure) to a 10 hPa (∼100 m) uniform vertical grid, then filtered using a the full profile using a fixed 40 hPa (∼400 m) filter window and quadratic fitting polynomial. We tested various combinations of filter windows and number of passes to determine the effects of filtering on the RS and RO profiles (not shown). The number of passes of the same filter result in minor smoothing effects relative to different filter windows, and filter windows larger than 40 hPa cause too much smoothing of the RS profile. RO profiles, due to their lower resolution relative to the RS profiles, showed little change under different filter windows, therefore, 40 hPa was chosen as a sufficient window to remove small-scale features in the RS profile while preserving the overall structures of both the RS and RO profiles. To accommodate for the differences in vertical resolution, RO profiles underwent a single pass of the 40 hPa filter and RS profiles underwent three passes of the 40 hPa filter.

2.3 RO–RS co-location criteria and statistics

We co-located RO and RS observations using the ellipse method every 10 hPa between 1000–10 hPa at each RS location. We included RO occultations taken within 3 h of the RS launch at Ny-Ålesund, Tateno, and Nauru and within 1 h at Lindenberg due to the high number of RS launches (at least four times daily). We considered three different geometric constraints to co-locate RO profiles: (1) large circle of radius 6 ° latitude (∼666 km), (2) small circle of radius 2.6 ° latitude (∼300 km), and (3) ellipse parallel to the wind direction, 6 ° latitude semi-major axis and 1.2 ° latitude (∼133 km) semi-minor axis, each centered at the RS location. (The ellipse and small circle are circumscribed by the large circle, and the small circle was chosen such that the area within the small circle and ellipse are approximately the same, see Fig. 2). The X and Y coordinates of the points on the circles and ellipses are constructed using the following parameterization (a: semi-major axis, b: semi-minor axis, for circles...
**Figure 2.** 500 hPa ERA-Interim refractivity field (color contours) centered at Tateno for Feb. 11 and 12, 2012 at 00:00 UTC. The three spatial geometries (large circle, small circle, ellipse) are shown in black centered at Tateno with wind vectors (black) denoting wind direction and magnitude. Generally, refractivity isopleths lie parallel to the wind flow, even as wind direction changes, resulting in minimal refractivity variability within the ellipse relative to both circles.

\[ a = b = \text{radius, and both } a, b \text{ are in km):} \]

\[
X(s) = a \cos(s) \cos(\theta) - b \sin(s) \sin(\theta) \\
Y(s) = a \cos(s) \sin(\theta) + b \sin(s) \cos(\theta)
\]  

(2)

where \( s \) is a parameter that varies between 0 and \( 2\pi \) and is stepped in increments of 0.01, yielding a series of 629 points that approximate the circle or ellipse, and \( \theta \) is the wind direction in radians (converted from meteorological to polar coordinates).

The \( X \) and \( Y \) coordinates of the circle or ellipse are first computed at the Equator where 1° of latitude and longitude equals 111 km, and then the \( X \) and \( Y \) coordinates are adjusted to the latitude and longitude of the RS station according to the following:

\[
\text{LAT}(Y, s) = \frac{Y(s)}{111} + \text{LAT(RS)} \\
\text{LON}(X, s) = \frac{X(s)}{111 \cos(\text{LAT}(Y, s))} + \text{LON(RS)}
\]  

(3)
The ellipses change orientation such that the semi-major axis is parallel to the wind direction at each pressure level and time per RS. The circles remain fixed and unaffected by the change in wind direction, pressure level, or time.

Figure 2 illustrates the three geometries centered at Tateno at 500 hPa for two days in February 2012, noting that the ellipse adjusts its orientation as wind direction changes with time and at a pressure level. Preliminary testing of the ellipse co-location method with ERA-Interim model refractivity fields confirmed that refractivity isopleths tend to follow wind flow (as illustrated in Fig. 2); thus orientation of the ellipse along the direction of the wind flow should increase atmospheric homogeneity relative to the large circle.

For each RS at a given time and pressure level, RO profiles are co-located with the RS profiles under the time and geometric constraints discussed above. There can (and often are) multiple co-location pairs with the same RS at a given time and pressure level. We computed differences for each co-location pair:

\[ X_{\text{RO-RS}} = X_{\text{RO}} - X_{\text{RS}} \]  

where \( X \) refers to time, refractivity, temperature, and/or water vapor pressure. We computed the RMS of the differences for each pressure level over the full time period and used the RMS to quantify the reduction in sampling errors. In certain cases we also computed the percent difference between two quantities \( X_i \) and \( X_j \):

\[ \% \text{diff} = 100 \times \frac{100 \times (X_i - X_j)}{X_j} \]  

2.4 **Spatial-temporal sampling** correction algorithm

**Figure 3.** Schematic of the spatial-temporal sampling correction applied to RO and RS co-location.
Applying a spatial-spatial-temporal sampling correction to the differences during RO–RS co-locations differences is an alternate method of reducing sampling errors in the presence of an auxiliary data set. This method has been applied in previous studies and is not restricted to RO–RS comparisons. Haimberger et al. (2012) used this approach to homogenize RS records. Wong et al. (2015) used ECMWF forecasts to estimate for double-differencing to reduce the sampling differences between Atmospheric Infrared Sounder (AIRS) and RS co-located pairs, using a double-difference estimate to correct AIRS and RS differences. Tradowsky et al. (2017) calculated an observed–background (O–B) double-difference to estimate the mean RS temperature bias using co-located RO profiles and the Met Office NWP background fields, in which the double-difference is applied to reduce the effects of temporal and spatial sampling errors. These studies use a double-difference sampling correction, but do not verify or discuss how the correction reduces sampling errors.

Here, we apply a spatial-spatial-temporal sampling correction double-difference computed with model data and assess its effects on reducing sampling errors. We use ERA-Interim data to subtract the model background from both the RO and RS observations, removing spatial and temporal sampling differences and isolating the observational errors associated with the RO and RS pair. As shown in Fig. 3, we use the interpolated values of the ERA-Interim profiles at the RO-interpolated in time and space to the RO time and location (eraPrf profiles provided by CDAAC) and use the ERA-Interim grid point nearest the RS location at the RS launch time to compute the spatial sampling correction. Due to the coarser vertical resolution of the ERA-Interim grid relative to the RS and RO vertical resolutions, comparisons with the spatial sampling correction are conducted on a common 50 hPa uniform pressure grid from 1000–100 hPa to avoid further vertical interpolation.

The spatial-spatial-temporal sampling corrected differences \( X^{sc} \) for the co-location pairs are computed as follows for the comparison variables (refractivity, temperature, and water vapor pressure):

\[
X^{sc} = \frac{(X_{RO} - X_{ERA(RO)}) - (X_{RS} - X_{ERA(RS)})}{(X_{RO} - X_{RS}) - (X_{ERA(RO)} - X_{ERA(RS)})}
\]

\[
X^{sc} = (X_{RO} - X_{ERA(RO)}) - (X_{RS} - X_{ERA(RS)})
\]

\[
= (X_{RO} - X_{RS}) - (X_{ERA(RO)} - X_{ERA(RS)})
\]

where \( X_{ERA(RO)} \) is the ERA-Interim model value interpolated in time and space to the RO location, \( X_{ERA(RS)} \) is the ERA-Interim model grid point value closest to the RS at the RS launch time, and their difference, \( X_{ERA(RO)} - X_{ERA(RS)} \), is referred to as the model correction term. We computed the RMS differences per pressure level on the 50 hPa grid.

The spatial-temporal corrected difference \( X^{sc} \) between the two data sets (Eq. (6) and (6a)) can be interpreted in two ways. When the two data sets are close to each other in time and space, one interpretation is that the RO–RS differences are corrected for local spatial and temporal variability by subtracting these sampling differences, which are estimated by another data set (in this case ERA-Interim) from the measured differences (Eq. (6a)). This is the interpretation used here and by Wong et al. (2015).

However, as the equivalent Eq. (6) shows, the corrected difference is also the difference of the departures of the two data sets from a common, reference data set, or the “double-differencing” method (Chander et al., 2013; Tradowsky et al., 2017). In this
interpretation, the two data sets are not necessarily required to be close in space or time; the method is valid as long as the biases of the reference data set do not vary over the spatial and temporal scales of the comparison. We tested the sensitivity of the sampling correction method to the spatial and temporal scale of the comparisons in Appendix A for RO–RS pairs over a much larger spatial scale (within circles of radius 15° latitude) and longer time window (24 h) at Lindenberg. We find that the RMS, mean, and standard deviations of the comparisons are insensitive over this range of spatial and temporal scales, allowing for a large increase in the number of co-located pairs of data.

3 Results and discussion

Preliminary proof-of-concept testing of the ellipse method using only ERA-Interim data demonstrated a significant reduction in RMS refractivity differences within the ellipse relative to both the large circle and circles of similar area to the ellipse (not shown here). In the following two sections, aspects of the ellipse co-location are analyzed at the Lindenberg station. The final section presents the results of both the ellipse co-location and spatial sampling correction at all four RS stations.

3.1 Filtered vs. unfiltered

Table 1. Unfiltered-Filtered vs. filtered-unfiltered RMS percent differences* for refractivity, temperature, and water vapor pressure at Lindenberg, 2014. RMS differences are computed using all co-located pairs within 100 hPa pressure layers for the unfiltered and filtered and unfiltered data sets, then the percent differences between the two are computed and reported here.

| Pressure (hPa) | Refractivity | | Temperature | | Water Vapor Pressure |
|----------------|--------------|----------------|----------------|----------------------|
|                | large circle | small circle | ellipse        | large circle | small circle | ellipse        | large circle | small circle | ellipse        |
| 100 – 200      | -2.94 %      | -6.87 %      | -5.96 %        | -3.32 %      | -7.99 %      | -6.55 %        | -0.68 %      | -0.56 %      | 0.50 %        |
| 200 – 300      | -1.28 %      | -3.90 %      | -2.34 %        | -1.30 %      | -4.12 %      | -2.46 %        | -0.78 %      | -1.39 %      | -1.65 %       |
| 300 – 400      | -0.34 %      | -0.90 %      | -1.03 %        | -0.20 %      | -0.81 %      | -0.71 %        | -0.60 %      | -0.53 %      | -1.02 %       |
| 400 – 500      | -0.80 %      | -2.00 %      | -1.22 %        | -0.22 %      | -0.76 %      | -0.49 %        | -1.03 %      | -2.48 %      | -1.32 %       |
| 500 – 600      | -0.80 %      | -1.87 %      | -1.23 %        | -0.30 %      | -0.64 %      | -0.49 %        | -0.83 %      | -2.17 %      | -1.38 %       |
| 600 – 700      | -1.05 %      | -1.53 %      | -1.36 %        | -0.33 %      | -0.65 %      | -0.65 %        | -1.03 %      | -1.74 %      | -1.59 %       |
| 700 – 800      | -0.89 %      | -1.51 %      | -1.21 %        | -0.26 %      | -0.61 %      | -0.65 %        | -0.84 %      | -1.54 %      | -1.18 %       |
| 800 – 900      | -0.97 %      | -1.54 %      | -1.12 %        | -0.32 %      | -0.57 %      | -0.21 %        | -0.85 %      | -1.44 %      | -0.99 %       |
| 900 – 1000     | -0.62 %      | -0.65 %      | -0.94 %        | -0.37 %      | -0.41 %      | -0.26 %        | -0.48 %      | -0.61 %      | -0.79 %       |

*percent differences computed as $100 \times \frac{(\text{RMS}_{\text{filt}} - \text{RMS}_{\text{unfilt}})}{\text{RMS}_{\text{unfilt}}}$

Filtering both the RO and RS profiles has a small, positive impact on reducing RMS differences in refractivity, temperature, and water vapor pressure. Compared to RMS differences computed using the unfiltered profiles at Lindenberg, filtering both the RO and RS profiles before co-location reduces RMS differences by about 1% on average, up to almost 8% in some instances (see Table 1). Within the large circle, filtering has less of an impact on reducing vertical representativeness percent errors since sampling errors tend to dominate (mostly due to large spatial differences). The RO–RS percent differences in the small circle
and ellipse are more affected by vertical representativeness errors since the sampling errors are relatively small. Based on these results, to reduce representativeness errors we filtered all RO and RS profiles before computing their differences to reduce representativeness errors.

3.2 Effects of wind speed on ellipse co-location

The relationship between the wind direction and horizontal variability (and sampling error) of refractivity is expected to break down for light wind speeds, and indeed we found that if wind speeds are low, the effectiveness of orienting the ellipse along the direction of wind flow is significantly reduced. We separated co-located RO–RS pairs at Lindenberg into two groups based on the reported wind speed of the RS at a given time and pressure level: (1) wind speeds less than 5 m s\(^{-1}\) and (2) wind speeds greater than or equal to 5 m s\(^{-1}\). Figure 4 shows the RMS differences in refractivity for each group.

As shown in Fig. 4, the clear distinction between RMS profiles for the ellipse and two circles when wind speeds are at wind speeds greater than 5 m s\(^{-1}\) essentially vanishes when wind speeds are less than 5 m s\(^{-1}\). Similarities in RMS refractivity differences when wind speeds are less than 5 m s\(^{-1}\) are likely associated with greater atmospheric homogeneity and weaker gradients in the region, resulting in similar observations within the area containing both circles and the ellipse. When wind speeds increase to larger than 5 m s\(^{-1}\), there is a greater separation between the circles and ellipse with respect to the RMS difference. Under moderate to high wind speed conditions, the ellipse reduces RMS refractivity differences relative to both the large and small circle, particularly below about 700 hPa. In the upper troposphere, both the ellipse and the smaller circle give similar results and smaller RMS differences than the large circle (Fig. 4). In the case of Lindenberg, since the majority of wind
speeds are greater than $5 \text{ m s}^{-1}$ (as noted by the differences in co-location counts under the two wind constraints in Fig. 4), low wind speeds do not have a significant effect on the overall RMS differences when co-location pairs are not separated by wind speed (see Fig. 5).

### 3.3 Ellipse co-location and spatial sampling correction applied at four RS stations

![Figure 5. RMS differences for RO–RS co-locations at Lindenberg, 2014. RMS differences with ellipse method only (solid; large circle: orange, small circle: grey, ellipse: magenta) and RMS differences with ellipse method and spatial sampling correction (dashed, abbreviated as SC) are shown for all three variables: (b): refractivity, (c): temperature, (d): water vapor pressure. The number of pairs within each ellipse (solid) and ellipse plus spatial sampling correction (dashed) are is shown in (a).](image)

We carried out two RO–RS comparisons at four different RS locations (Lindenberg, Ny-Ålesund, Tateno, and Nauru): first we compared pairs with RO observations co-located within the large circle, small circle, and ellipse centered at the RS station, and second, we applied the spatial sampling correction to the RO–RS pairs within the ellipse and two circles. We then computed RMS differences in refractivity, temperature, and water vapor pressure.

Figure 5 illustrates the RMS differences for refractivity, temperature, and water vapor pressure with and without the spatial sampling correction at Lindenberg. Considering the ellipse method only (solid lines), the ellipse reduces RMS differences relative to the large circle for all three variables at all pressure levels, having the most significant reduction in the temperature RMS differences. Generally, the small circle and ellipse have similar RMS differences, however there are pressure layers in which the ellipse reduces the RMS relative to the small circle. When the spatial sampling correction is applied to the circles and ellipse (dashed lines), the RMS differences are significantly reduced and converge with minimal differences between the ellipse and circles.

The RMS differences at Lindenberg are mostly affected by spatial sampling errors; temporal sampling errors are minimized by the reduced time window of 1 h due to frequent (four times daily) RS launches at this station. The ellipse decreases RMS
differences relative to both circles, effectively reducing spatial sampling errors driven by atmospheric variability. The spatial correction is most effective at reducing spatial sampling errors, particularly for the large circle in which atmospheric variability is more likely.

Figure 6. Same as Fig. 5 for two different locations: Ny-Ålesund, 2013 (a-d) and Tateno, 2012 (e-h).

The results at Ny-Ålesund and Tateno are very similar to those at Lindenberg, with some minor differences in the lower troposphere (Fig. 6). Again, using pairs in the ellipse only (solid, magenta) reduces results in significantly reduced RMS differences relative to the large circle (solid, orange) and is equal to or less than RMS differences within smaller RMS differences compared to the small circle (solid, grey). Including the spatial sampling correction shows the largest reduction in RMS differences such that the differences from all three geometric types converge to similar values. At both Ny-Ålesund and Tateno, the reduction of RMS differences in the ellipse compared to the large circle becomes less in the lower troposphere (below 800 hPa) where the frequency of wind speeds less than 5 m s\(^{-1}\) increases and the relationship between wind direction and RO–RS differences breaks down. For example, at Ny-Ålesund and Tateno the percentages of RO–RS pairs for which
the wind speed is less than 5 m s\(^{-1}\) below 800 hPa are 36.1% and 32.7% respectively. The small sample size in the ellipse compared to the large circle (Fig. 6) may also play a role, allowing a few outliers to dominate the statistics.

**Figure 7.** Same as Fig. 5 for Nauru, 2011-2013.

The results at the tropical location of Nauru generally confirm the findings at the other three RS locations. The ellipse alone (Fig. 7, solid, magenta) decreases RMS differences relative to the large circle (solid, orange) and is produces RMS differences that are comparable to the RMS-differences of the small circle (solid, grey). There is less distinction, however, between the ellipse method only (solid) and spatial sampling correction (dashed) RMS differences, particularly for temperature. All six geometric and spatial sampling correction combinations have overlapping RMS temperature differences and show little separation between co-location methods, unlike the results at Lindenberg, Ny-Ålesund, and Tateno. This is most likely caused by the relative horizontal homogeneity in temperature at Nauru, which is located in the deep tropics. Thus, there is little to no distinction between RMS temperature differences using various geometrical constraints or spatial sampling correction.

Refractivity and water vapor pressure RMS differences in the lower troposphere (1000 – 700 hPa) at Nauru are the largest of the four stations, caused primarily by atmospheric conditions at its location in the deep tropics. RO is known to have negative refractivity biases in the lower troposphere, particularly in moist tropical regions where large water vapor and associated refractivity gradients often result in super-refraction (Rocken et al., 1997; Ao et al., 2003; Sokolovskiy, 2003; Beyerle et al., 2006; Anthes et al., 2008). Overall, this results in larger RMS refractivity differences at Nauru. Since super-refraction is an error characteristic of RO retrievals and not related to horizontal sampling errors, the spatial sampling correction has no impact on the RO–RS refractivity differences.
We have shown that vertical filtering of the RO and RS profiles before comparison reduces representativeness errors associated with different vertical resolutions and observation types by a small amount (typically a few percent). Using these filtered profiles, we tested two methods to reduce spatial and temporal sampling errors during RO–RS co-location comparisons: (1) restricting RO and RS pairs to within ellipses oriented along the direction of wind flow and (2) applying a spatial-temporal sampling correction using model data to remove differences caused by horizontal atmospheric gradients and time differences in the observations. When wind speeds exceed about 5 m s\(^{-1}\), co-locations within the ellipse parallel to the wind flow reduce RMS differences in refractivity, temperature, and water vapor pressure relative to co-locations within the large circle, and either reduce or are equal to RMS result in RMS differences that are approximately equal to the differences within the smaller circle.

The effectiveness of co-locating RO–RS pairs within the ellipse is reduced for wind speeds less than 5 m s\(^{-1}\).

Applying the spatial-temporal sampling correction using ERA-Interim model data showed the most significant reduction in RMS differences, more-so than applying the ellipse constraint alone. The spatial sampling correction reduced RMS differences in refractivity, temperature, and water vapor pressure by an average of 55%. The reductions of sampling errors within both large and small circles and the ellipse tend to converge with the spatial sampling correction applied, rendering the differences in geometric constraints of the circles and ellipse negligible. An exception to this reduction in RMS occurs at Nauru in the lower troposphere, where super-refraction associated with the atmospheric conditions of the deep tropics tends to dominate the RMS differences.

In order to reduce sampling errors for future RO–RS co-location comparisons, our results suggest that applying the spatial sampling correction under more lenient co-location criteria would be most effective. By using a large distance constraint, the sample size will be sufficiently large and applying the spatial sampling correction eliminates most sampling errors, even for the large distance restriction (greater than 600 km). However, if a reliable model is unavailable, restricting co-locations within ellipses oriented along the direction of wind flow will help to reduce sampling errors caused by atmospheric variability. Both of these methods are effective in reducing sampling errors caused by spatial and temporal differences during comparisons and should provide a more accurate error analysis of RO and RS observations.

Code availability. The code used in this study will be made available upon request.

Data availability. The data used in this study will be made available upon request.

Appendix A

A1 Mean and standard deviation of differences
In this section, we compute the mean and standard deviation (STD) of the RO–RS differences. The mean difference is defined as

\[
M(RO - RS) = \frac{1}{N} \sum_{i=1}^{N} (RO - RS)_i
\]  

(A1)

The variance about the mean is given by

\[
VAR_M(RO - RS) = \frac{1}{N} \sum_{i=1}^{N} [(RO - RS)_i - M(RO - RS)]^2
\]  

(A2)

and the standard deviation of the difference about the mean is the square-root of the variance, \(\text{STD}_M = \text{VAR}_M^{1/2}\). The \(\text{STD}_M\) can be compared to the RMS RO–RS difference, which is defined by

\[
\text{RMS}(RO - RS) = \left[ \text{MS}(RO - RS) \right]^{1/2}
\]  

(A3)

where the mean square difference (MS) is defined as

\[
\text{MS}(RO - RS) = \frac{1}{N} \sum_{i=1}^{N} (RO - RS)_i^2
\]  

(A4)

It can be shown by expanding Eq. (A2) that

\[
\text{VAR}_M(RO - RS) = \text{MS}(RO - RS) - [M(RO - RS)]^2
\]  

(A5)

Thus, the \(\text{STD}_M(RO - RS)\) is always less than or equal to the RMS(RO – RS), and equal if and only if the mean RO–RS difference is zero.

Figures A1–A3 show the mean and STD of the RO–RS differences for refractivity, temperature, and water vapor pressure, respectively, for all six co-location methods. For each variable, the mean difference profiles are all similar and show no systematic differences, which indicates that the mean differences are not sensitive to the co-location method and are relatively unaffected by sampling errors. This is because the distribution of the RO observations with respect to the RS observations is approximately random, and so the sampling (and representativeness) errors tend to cancel for a large enough sample size.

However, the mean differences do illustrate differences in bias errors between the two data sets. The bias errors in the GRUAN RS are expected to be small, so the differences in the mean RO–RS differences are likely due to the RO biases, notably the negative refractivity bias caused by super-refraction in the lower troposphere under moist conditions.

In Fig. A1 we see that the differences in mean refractivity are small above 700 hPa for all four stations. Below 700 hPa, there is a negative RO refractivity bias, which is most pronounced at the tropical Nauru station where super-refraction is most common. The bias is minimal at the most northern station, Ny-Ålesund, which is expected to have the fewest cases of super-refraction.
Figure A1. RO–RS refractivity mean (thick) and STD (thin) difference profiles for the six co-location methods at Lindenberg (a), Ny-Ålesund (b), Tateno (c), and Nauru (d). The number of co-located RO–RS pairs are given on the left side of each panel (without the sampling correction: solid; with the sampling correction: dashed). The color schemes are the same as Figs. 4–7 and are given in the legend in panel (a) where SC stands for sampling correction.

The biases for temperature differences (Fig. A2) are less than 1 K from 1000 to 100 hPa for Lindenberg, Ny-Ålesund, and Tateno, and show a negative bias at Nauru reaching about 2 K at 600 hPa. This is probably not caused by super-refraction in the RO observations because the negative bias is above the level where most super-refraction occurs and the bias becomes smaller below the 600 hPa level, in contrast to the bias in refractivity which increases sharply with decreasing height in the
lower troposphere. Instead, the bias in the temperature is probably related to the 1D-VAR retrieval of temperature in the RO observation, which uses ERA-Interim temperatures as the first guess.

Finally, the differences in the mean RO–RS water vapor pressure reflect the negative refractivity bias due to super-refraction at Lindenberg, Tateno, and especially Nauru, with little bias at the colder, drier Ny-Ålesund station (Fig. A3).

The STD profiles in Figs. A1–A3 are similar to the RMS profiles associated with the different co-location methods in Figs. 5–7. This is expected because the STD and RMS values are close when mean RO–RS differences are near zero, as shown by Eq. (A5). The largest differences between the STD and RMS profiles are in the lower troposphere, where the mean differences are the largest.
Figure A3. Same as Fig. A1 for water vapor pressure.

A2 Sensitivity of spatial-temporal correction method to spatial and temporal scale

In Sect. 2.4, we noted that when using the double-differencing correction method to reduce spatial and temporal sampling errors, it was not necessary that the RO profiles be close in space and time to the RS profiles, as long as the biases of the reference data set (in our case ERA-Interim) remain constant in space and time throughout the comparison. To illustrate this property of the double-differencing method, we co-located RO profiles within a much larger circle (15° latitude, ~1665 km) and a time window of 24 h at Lindenberg between January and March, 2014, and compared the RMS, mean, and STD of these RO–RS differences to those computed under the 6° and 2.6° circles with a 1 h time window.
Figure A4. RMS (b-d), mean (thick; e-g), and STD (thin; e-g) RO–RS profiles for refractivity (b, e), temperature (c, f), and water vapor pressure (d, g) under three different co-location restraints with (dashed) and without (solid) the sampling correction (SC) applied: circles of radius 6° latitude (orange) and 2.6° latitude (grey) for time windows of 1 h, and a circle of radius 15° latitude for a 24 h time window (black). Comparisons are conducted at Lindenberg for January–March, 2014. The number of co-located pairs for each co-location restraint (with SC: dashed; without SC: solid) are given in panel (a), where the x-axis is a logarithm scale.

Figure A4 illustrates the RMS (b-d), mean, and STD (e-g) profiles for the RO–RS differences in refractivity, temperature, and water vapor pressure under the three different co-location restraints with and without the sampling correction applied. Across all three variables, the RMS, mean, and STD profiles with the sampling correction remain nearly the same regardless of the spatial or temporal differences applied to the RO–RS co-location. This verifies that the double-differencing method is insensitive to spatial and temporal separations of the RO and RS observations for this example when using ERA-Interim as a reference data set. The similarity in the RMS, mean, and STD profiles with the sampling correction also indicates that any existing bias in the ERA-Interim reference data set is nearly constant over these spatial and temporal scales.
The above results show that using the double-differencing method to reduce spatial and temporal sampling errors in RO–RS comparisons allows for many more RO–RS pairs to be included in the comparison (more than 35,000 for the $15^\circ$ with 24 h time window compared to approximately 250 for the $6^\circ$ circle and 1 h time window), as illustrated Fig. A4a, provided that the bias of the reference data set does not vary significantly over the spatial and temporal scales of the comparison. Increasing the size of the spatial and temporal scales of the comparison is thus a tradeoff between reducing the random error effects by increasing sampling size and possibly increasing the errors by allowing a greater effect of varying biases in the reference data set.

**Author contributions.** All co-authors contributed to developing the ideas and methodologies of this project. SG conducted the majority of the coding, data retrieval, computations, and analysis, with TR contributing to coding and data retrieval. SG prepared the manuscript with contributions from both co-authors.

**Competing interests.** The authors declare that they have no conflict of interest.

**Acknowledgements.** The authors thank Drs. Eric DeWeaver (NSF) and Jack Kaye (NASA) for their support of this research through NSF-NASA grant AGS-1522830. Dr. Sergey Sokolovskiy provided many useful comments and suggestions during this study. The first author was supported in part by the Significant Opportunities in Atmospheric Research and Science (SOARS) program, NSF grant AGS-1641177, and by the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) program at UCAR, which is sponsored by the National Space Office in Taiwan, NSF, NASA, NOAA, and the U.S. Air Force. Thanks to CDAAC for the provided RO data sets, and NOAA NCDC and GRUAN for the provided RS data sets. The authors thank Dr. Keith Maull for his suggestions during this study. The authors would like to thank the three anonymous reviewers for their comments and constructive suggestions which improved this manuscript.
References


